PHYSICAL ACTIVITY RECOMMENDATION FOR DIABETIC PATIENTS BASED ON ONTOLOGY

Djoko Pramono[1], **Nanang Yudi Setiawan**[2], **Riyanarto Sarno**[3], **Mohamad Sidiq**[4]
Department of Informatics Engineering, Faculty of Information Technology
Institut Teknologi Sepuluh Nopember Surabaya Indonesia 60111
[1]djoko.pramono12@mhs.if.its.ac.id, [2]nanang.yudi12@mhs.if.its.ac.id,
[3]riyanarto@if.its.ac.id, [4]sidiq12@mhs.if.its.ac.id

ABSTRACT

A recommendation system can help people with better decision making from much complex information. This paper presents a recommendation system in physical activity/exercise during medical treatments for a diabetic patient. Physical activity is part of the management of diabetes mellitus which can improve glucose level by increasing insulin sensitivity and keeping carbohydrate metabolism (body weight loss). When combined with diet and drug therapy, physical activity contributes significantly in the improvement of glycaemic control. Diabetes care is a complex medication, dealing with uncertainty issues. To accommodate this uncertainty and vagueness, a system based on ontology introduced. It provides standard vocabulary, technical terminology and a domain model for knowledge integration. We introduce a recommendation system based on ontology to determine the physical activity guidance according to patient medical evaluations including ages, complications, daily activities, and food intakes. First, we build the ontology of diabetic knowledge refer to American Diabetes Association statements. Then, use similarity algorithm, *weighted Directed Acyclic Graph* (wDAG), to find semantic relationship and interoperability among patient ontology to classify the diabetic patient condition. Therefore, system will able to choose the most appropriate physical activities for this patient.

Keywords: Diabetes application, Ontology, weighted Directed Acyclic Graph.

1 INTRODUCTION

At certain level, physical activity contributes considerable on healing or treatment of diabetes mellitus patients. Physical activities as exercises practiced regularly and correctly, reduce cardiovascular risk factors, contribute to weight loss, and improve well-being [1].

.

Anti-diabetic drug recommendations also have the same characteristics, where a drug type or prescription for every patient would differ depending on the patient's current condition. Both of these diabetes medical treatments have a common complexity and uncertainty problems. Research by Chen[2], resulting a recommendation system of anti-diabetic drugs selection based on ontology. In the evaluation test, he proved that his research for anti-diabetic drugs recommendation system met the doctors need. He successfully developed a system with precision rate up to 100% on 20 patients tested. An ontology based systems proved to be applied for the same characteristic of uncertainty, as in the selection of diabetic prescriptions depending on the patient's conditions.

An approach in tree or graph algorithm has been widely used to resolve problems of ontology. Aranguren[3] used a *Directed Acyclic Graph* (DAG) for recasting the Gene Ontology. Thus, the DAG representation also be used for matchmaking or measuring the similarity of two ontology. If the degree of importance between nodes presents, a weight on the arcs placed. Research by Jin[4] comes with an algorithm for computing the simplicity of node-labeled, arc-labeled and arc-weighted DAGs, and the similarity between pairs of such wDAGs (weighted Directed Acyclic Graphs). In SOA development, Sarno[5] used the wDAG to develop the schemas for semantic web service and to compute the similarities applied for semantic matchmaking.

This paper proposes a recommendation system based on ontology that provides physical activity/exercise recommendations for diabetic patients. The recommendation will suggest types of exercises including its intensity, frequency, and duration in accordance with patient's conditions (age, complication, body mass index, calorie consumption, type of diabetes). In our study, a patient can have a different input attributes than patients in the system (or different ontology). Several arcs having different weights, which describe the different relative importance of the corresponding arcs to the attribute, can link this attribute.

2 MODEL, ANALISIS, DESIGN, AND IMPLEMENTATION

2.1 Ontology

Gruber [6] defines ontology as "*An ontology is an explicit specification of a conceptualization*". The ontology described as a union of conceptualization and systematic description of the unit. Ontology is used to express explicit and formal skills through conceptual structures [7]. The ontology provides a shared dictionary, which can be used to model a domain, under which, the type of objects and/or concepts are available, the following relation and its properties [8].

An ontology generally contains a list of limited words (terms) and relationship between words. This term indicates the main concept (object class) of a domain, while its boundary includes hierarchy of classes. Ontology should include other information, such as property, value, disjointedness, and the specification of the logical attachment between objects. Composition of ontology is a knowledge-intensive approach, and may be treated as one form of an engineering knowledge, involving multiple acquisition, modeling, and process knowledge representation in a row [9][10]. The latest research on the development of knowledge-based systems showed the applications of ontology technology still arising in several domains.

The main task in the preparation of the ontology is to translate the goal-oriented activity or troubleshooting a systematic knowledge to solve problems. Uschold and Grueninger[11] observed that the type or domain issues affecting the needs of knowledge in solving. Ontological classification showed a different knowledge needed in each class.
Ontology consist concept, relationship, and *instances*, that as following: [12]

1. *Concept* or *Class*: a concept suggests topics or something in the domain ontology.
2. *Relationship* or *Attribute*: The relationship is correlation between concepts when considering concept. In other words, relationship is extension of concepts.
3. *Instance*: The instance is revealed by a series of concepts and relationships, which have specific knowledge such as web pages, documents, and so on.

Thus, it can be represented with an OWL ontology given the OWL contains classes, properties and individuals, which roughly correspond to the concepts, rules and facts in the ontology.

2.2 Weighted Directed Acyclic Graph (WDAG)

To measure the degree of relative between weights on different attributes, Jin[4] introduce wDAG. WDAG was the DAG representation with arc-labeled, arc-weighted, i.e.:"*An arc-labeled, arc-weighted DAG is constructed from a 6-tuple (V, E, L_V, L_E, L_W, r) of a set of nodes V, a set of arcs E, a set of arc labels L_V, and a set of arc labels L_E, a set of arc weights L_W= [0,1], and one element $r \in V$, respectively, such that (V, E, L_V, L_E, r) satisfies the definition of an arc-labeled DAG and there is a many-to-one mapping from the elements in E to the elements in L_W (i.e. different arcs can carry the same weights).*"

WDAG similarity algorithm compares and calculates the similarity between two arc-labeled, arc-weighted DAG. The application of this algorithm by incorporating both wDAG serialization in XML using OO RuleML (Object-Oriented Rule Mark-up Language), then read by the wDAG similarity algorithm. The output of this algorithm is the value of the similarity between both wDAG. Figure 1 shows flow diagram of wDAG similarity algorithm.

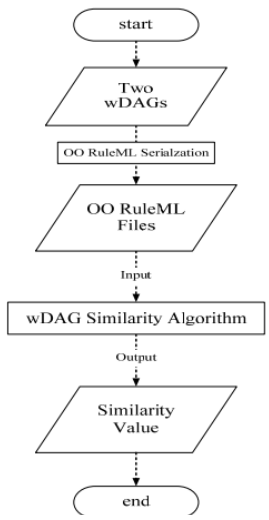

Figure 1. Flow chart of the wDAG similarity algorithm [4]

The main function *wDAGsim*(*g, g'*), computes the similarity of two (sub)wDAGs, which is formulated in equation (1).

$$
wDAGsim(g,\ g') = \begin{cases} 0.0 & \textit{the root node labels of g and g' are not identical} \\ 1.0 & \textit{g and g' are leaf nodes} \\ \sum \begin{cases} wDAGsim(g_i, g'_j)\cdot\dfrac{(w_i + w'_j)}{2} & g_i \textit{ and } g'_j \textit{ are not missing} \\ wDAGsim(g_i, \varepsilon)\cdot\dfrac{(w_i + 0)}{2} & g_i \textit{ is missing in } g' \\ wDAGsim(\varepsilon, g'_j)\cdot\dfrac{(0 + w'_j)}{2} & g'_j \textit{ is missing in } g \end{cases} \\ \sum_{j=1}^{breadth\_of\_g'} wDAGsim(\varepsilon, g'_j)\cdot\dfrac{(0 + w'_j)}{2} & \textit{only g is a leaf node} \\ \sum_{i=1}^{breadth\_of\_g} wDAGsim(g_i, \varepsilon)\cdot\dfrac{(w_i + 0)}{2} & \textit{only g' is a leaf node} \end{cases} \quad (1)
$$

where:

wDAGsim(g, g'): similarity of two input wDAGs g and g'.

wDAGsim(g_i, g'_j): intermediate similarity of the i[th] and j[th] sub-wDAGs of the wDAGs g and g', respectively.

w_i and w'_j: arc weights of the i[th] and j[th] child of the wDAG g and g', respectively.

i: increase from 1 to the breadth of g.

j: increase from 1 to the breadth of g'.

ε: an empty wDAG.

The wDAG similarity algorithm traverses recursively using depthfirst strategy into the two wDAGs g and g' and computes their similarity form bottom to upper layer (bottom-up). If a sub-wDAG of g or g' is unavailable, then computes the simplicity of two (sub)wDAGs shown at equation (2).

$$
wDAGplicity(g) = \begin{cases} D_I & \textit{if g is a leaf node} \\ \dfrac{D_F}{m}\sum_{j=1}^{m} w_j \cdot wDAGplicity(g_j) & \textit{otherwise} \end{cases} \quad (2)
$$

where,

D_I and D_F: depth degradation index and depth degradation factor.

m: breadth of the wDAG g that is not a leaf.

The simplicity is computed recursively through top-down traversal into the sub-wDAGs. The simplicity value of a wDAG g is the sum of the simplicity values of its sub-wDAGs multiplied by the arc weights, a sub-wDAG depth degradation factor, and a sub-wDAG depth degradation index [5].

2.3 Recommendation System

System endorses the techniques and software that can provide the consideration/advice an item, which should be used by users [13]. Customization provided by the system, focused on providing advice that is useful and effective against a specific type of item

2.4 Knowledge Construction

The physical activity recommendation system designed based on knowledge about the kinds of physical activity that correspond to the conditions of diabetic patients. The knowledge is based on the joint position statement of Physical Activity/ Exercise and Diabetes Mellitus [1]. This knowledge formulated in two schemas:

1. Ontology of patients with these types of patients with individuals that are a combination of the various parameters of the types of diabetes, age, BMI, blood sugar levels and suffered complications.
2. Ontology of physical activity with these kinds of sports with individuals that a combination of exercise, intensity, frequency, and duration.

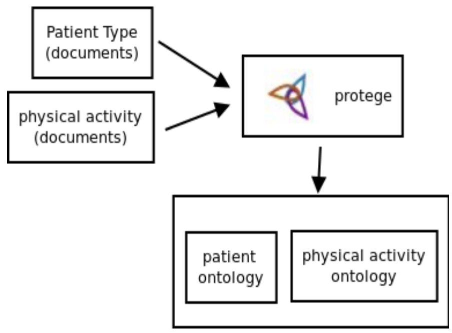

Figure 2. Processing steps of ontology building

Protege is used to compile the data already collected into the ontology. Protege produces ontology in OWL or RDF format. Protege support SPARQL to query concepts in the ontology. With known individual from the patient's ontology we can query taking individuals on the ontology of physical activities that have levels of calories must be burned, the duration and the intensity of the corresponding to the selected individual patient.

Our study uses the ontology similarity algorithm to determine the type of individuals that fit patient based on patient's data. Methods in this study, based on ontology similarity algorithm which produces the value weights from the results of a comparison between evaluated patient's models compared to models generated from patients ontology in the system knowledge base.

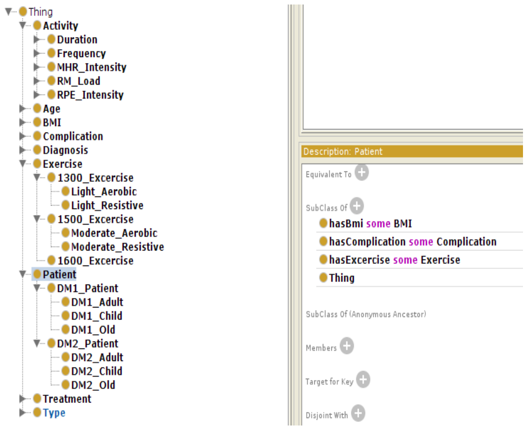

Figure 3. Patient ontology and exercise ontology

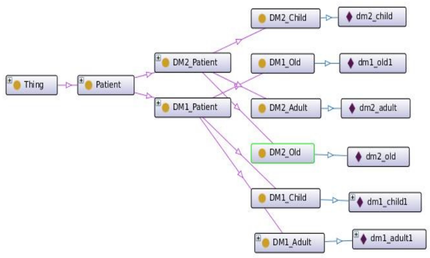

Figure 4. Individuals in patient ontology

2.5 Ontology Similarity

WDAG similarity calculates similarity value between evaluated patients with individual patients ontology owned by the system (from medical references). User lists the patient's data through the application interface. Its contain type of diabetes, age, type of work activities, body weight and height, and complication diseases.

We will have an ontology *p* representing the structure or schema of input data. Our knowledge system stores a wide range of patients as an individual ontology *x* for a single record. So there will be x_1,x_2,... and x_n for n^{th}-patient. Once formed, ontology *p* would be compared across the individuals (x_1,x_2,...,x_n) with wDAG similarity algorithm. It comes up an individual *x* with the highest similarity value. Selected individual will be the key or constraint value in data retrieval query using SPARQL. Brought a similar patient schema with physical activities recommendation that satisfies the evaluated diabetic patient to improve insulin sensitivity and assist in diminishing elevated blood glucose levels into the normal range.

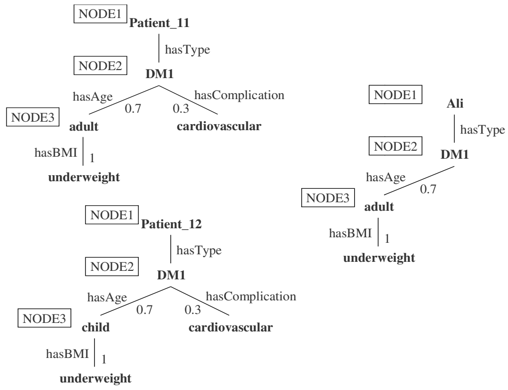

Figure 5. Ontology similarities between patients

Figure 5 shows Patient_11 and Patient_12 ontology built from physical activity guidance. Ali's ontology becomes a patient which is system willing to recommend. Labels put to every node on both ontology compared. For the same node in both trees, has the same label. Next, we use wDAG similarity formula (both equation 1 and 2) to get the highest similarity value of all patient ontology. Table 1 and 2 describe the similarity calculation stages at tree nodes between patient ontology. Simplicity value shows up when *hasComplication* property is not available at Ali's ontology. The similarity calculations are given below:

$$sim(P11.N3, A.N3) = \frac{1+1}{2} \times 1 = 1$$

$$sim(P11.N2, A.N2) = \frac{0.7+0.7}{2} \times 1 = 0.7$$

$$plicity(P11.N2) = 0.5 \times 0.3 = 0.15$$

$$\begin{aligned} sim(P11.N2, A.N2) &= 0.7 + (0.15 \times 0.5) \\ &= 0.7 + 0.075 \\ &= 0.775 \end{aligned}$$

$$\begin{aligned} sim(P11.N1, A.N1) &= \frac{1+1}{2} \times 0.775 \\ &= \mathbf{0.775} \end{aligned}$$

Table 1. Similarity between patient_11 and Ali

		A.wDAG		
		A.NODE1	A.NODE2	A.NODE3
P11.wDAG	P11.NODE1	sim(P11.Nd1, A.Nd1)		
	P11.NODE2		sim(P11.Nd2, A.Nd2)	
	P11.NODE3			sim(P11.Nd3, A.Nd3)

		A.wDAG		
		NODE1	NODE2	NODE3
P11.wDAG	NODE1	0.775		
	NODE2		0.775	
	NODE3			1

	A.wDAG		
	NODE1	NODE2	NODE3
Simplicity			0.15

$$plicity(P12.N3) = 0.5 \times 1 = 0.5$$

$$\begin{aligned} plicity(P12.N2) &= 0.5 \times 0.5 \times ((0.5 \times 0.7) + 0.3) \\ &= 0.25 \times 0.65 \\ &= 0.1625 \end{aligned}$$

$$\begin{aligned} sim(P12.N2, null) &= 0.5 \times 0.1625 \\ &= 0.08125 \end{aligned}$$

$$plicity(A.N3) = 0.5 \times 1 = 0.5$$

$$\begin{aligned} plicity(A.N2) &= 0.5 \times 0.5 \times 0.7 \\ &= 0.175 \end{aligned}$$

$$\begin{aligned} sim(A.N2, null) &= 0.5 \times 0.175 \\ &= 0.0875 \end{aligned}$$

$$\begin{aligned} sim(P12.N2, A.N2) &= 0.08125 + 0.0875 \\ &= 0.16875 \end{aligned}$$

$$\begin{aligned} sim(P12.N1, A.N1) &= \frac{1+1}{2} \times 0.16875 \\ &= \mathbf{0.16875} \end{aligned}$$

Table 2. Similarity between patient_12 and Ali

		A.wDAG		
		A.NODE1	A.NODE2	A.NODE3
P12.wDAG	P11.NODE1	sim(P12.Nd1, A.Nd1)		
	P11.NODE2		sim(P12.Nd2, A.Nd2)	
	P11.NODE3			

		A.wDAG		
		NODE1	NODE2	NODE3
P12.wDAG	NODE1	0.16875		
	NODE2		0.16875	
	NODE3			

	A.wDAG		
	NODE1	NODE2	NODE3
Simplicity		0.175	0.5

	P12.wDAG		
	NODE1	NODE2	NODE3
Simplicity		0.1625	0.5

From the results obtained by calculation that the patient named Ali has the similarities value to the individual Pateint_11 of 0,775 and similarities to individual Pateint_12 of 0,16875. The selected individual Pateint_11 will be a reference to retrieve the physical activity recommendations for similar tested patients.

Using SPARQL query shown at Figure 6, those recommendations could be retrieved from our knowledge system based ontology. The query is integrated in the platform-independent, object-oriented Java application programming for user comfortableness and operability.

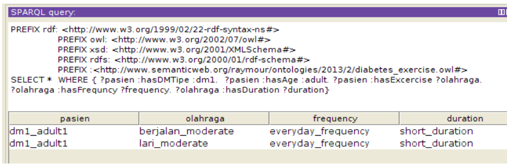

Figure 6. SPARQL query for patient physical activity

3 EVALUATION

Our recommendation system for physical activity on diabetic patient is an ongoing developed system and no formal evaluations have yet taken place. We had planned the system evaluations for several diabetic patients with varying conditions on ages, diabetes mellitus type, body mass index, and complications. The results will show the performance of the proposed system generates the same prescription released by an expert (or a doctor). We prepared a form containing list of patients with their conditions and physical activities recommendation from the system. An expert clarifies our recommendation by fulfilling the available column according to his expertise. Table 3 shows the evaluation form for our system assessed by an expert (doctor or medical therapist). We used Cohen's Kappa statistic for measuring agreement between system and doctor assessment. The kappa coefficient is an appropriate measure of reliability for data evaluation lie on a nominal or an ordinal scale [15].

Table 3. Evaluation form for the recommendation system

<table>
<tr><th>ID</th><th>Condition</th><th>Data</th><th>System</th><th>Expert</th></tr>
<tr><td rowspan="14">Patient1</td><td>Age</td><td>Adult</td><td rowspan="14">Activity: ....
Intensity: ....
Duration: ....
Frequency: ....</td><td rowspan="14">Agree/ Not</td></tr>
<tr><td>DM Type</td><td>2</td></tr>
<tr><td>Complications</td><td></td></tr>
<tr><td>Cardiovascular</td><td></td></tr>
<tr><td>Peripheral arterial</td><td>yes</td></tr>
<tr><td>Retinopathy</td><td>yes</td></tr>
<tr><td>Nephropathy</td><td></td></tr>
<tr><td>Peripheral neuropathy</td><td></td></tr>
<tr><td>Autonomic neuropathy</td><td></td></tr>
<tr><td>BMI</td><td>Overweight (26)</td></tr>
<tr><td>Weight</td><td>95</td></tr>
<tr><td>Height</td><td>69</td></tr>
<tr><td>Blood Glucose</td><td>9 mmol/l</td></tr>
</table>

Cohen's kappa measures the agreement between two raters who each classify N items into C

mutually exclusive categories. Equation (3) shows Cohen's kappa coefficient equation.

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where:

$\Pr(a)$ is the relative observed agreement among raters, and $\Pr(e)$ is the hypothetical probability of chance agreement. If the raters are in complete agreement then K = 1. If there is no agreement among raters then K = 0.

4 CONCLUSION

There are many researches on anti-diabetic drug recommendations, for doctors to prescribe, few based on ontology. First, our study used Protégé to build the patient and physical activity ontology knowledge. Next, wDAG was used to get degree of similarity between patient ontology. Finally, Semantic searching through the system is made by SPARQL, and it carried out the physical activity recommendation about instances of type, intensity, frequency and duration of physical activity. Diabetes medical care guidance was imported into the knowledge system. The guidance we used, "Physical Activity/Exercise and Diabetes Mellitus" is a joint position statement between the American Diabetes Association and the American College of Sports Medicine.

Our system will choose the best physical activity guideline considering patient condition and previous medical examination. Though we still have problems in getting medical diabetic patients especially with their physical therapy, based on previous research, we evaluate the system with diabetic patient corresponding data. In future work, we will strengthen patient ontology and evaluate more patient data to our system. In combination with dietary (food intake), hopefully we will have a better recommendation taking into account the number of calories obtained from food consumption and issued through physical activity.

REFERENCE

[1] ADA (American Diabetes Association), Standards of Medical Care in Diabetes, *Diabetes Care*, Volume 36, Supplement 1, 2013, 73-77.

[2] Rung-Ching Chen, Yun-Hou Huang, Cho-Tsan Bau, and Shyi-Ming Chen, A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection, *Expert Systems with Applications*, 39, 2012, 3995–4006.

[3] Mikel Egana Aranguren, Sean Bechhofer, Phillip Lord, Ulrike Sattler, and Robert Stevens, Understanding and using the meaning of statements in a bio-ontology: recasting the Gene Ontology in OWL, *BMC Bioinformatics*, 2007.

[4] Jing Jin, Similarity of weighted directed acyclic graph, *Master's thesis University of New Brunswick*, 2006.

[5] Riyanarto Sarno, Khakim Ghozali, Bayu Adhi Nugroho, and Astria Hijriani, Semantic Matchmaking using Weighted Directed Acyclic Graph. Surabaya, *3rd Applied Technology Science and Art*, 2011, 1-6.

[6] T. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5(2), 1993, 199–220.

[7] Chi, Y. L., Ontology-based curriculum content sequencing system with semantic rules, *Expert Systems with Applications*, 36(4), 2009, 7838–7847.

[8] Fredrik Arvidsson and Annika Flycht-Eriksson, Ontologies I, *Homepage of PhD Course - Semantic Web 2002*, Retrieved 6 April 2013.

[9] N. Guarino, and P. Giaretta, Ontologies and knowledge bases: Toward a terminological clarification. *In N.J.I. Mars (ed.), Towards Very Large Knowledge Bases*, IOS Press, 1995, 25-32.

[10] N. F. Noy and C. D. Hafner, The state of the art in ontology design, *AI Magazine*, 18(3), 1997, 53–74.

[11] M. Uschold, and M. Grueninger, Ontologies: Principles, methods and applications, *The Knowledge Engineering Review*, 11(2), 1996, 93–155.

[12] Rung-Ching Chen, Pao-Cho Tsan, I-Yen Lee, and Jiin-Chyr Hsu, Medical Adverse Events Classification for Domain Knowledge Extraction, Ninth International Conference on Hybrid Intelligent Systems, 2009, 298-303.

[13] F. Ricci et al., *Recommender Systems Handbook*, DOI 10.1007/978-0-387-85820 - 3_1, Springer ScienceBusiness Media, LLC, 2011.

[14] Prem Melville and Vikas Sindhwani, *Encyclopedia of Machine Learning*, Springer-Verlag Berlin Heidelberg, 2010.

[15] Julius Sim, The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements, *Physical Therapy*, 85, 2005, 257-268.